



TECH CENTER 1600/2900

1600

## RAW SEQUENCE LISTING

DATE: 02/20/2003

PATENT APPLICATION: US/09/765,272A

TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

## SEQUENCE LISTING

## 5 (1) GENERAL INFORMATION:

- 7 (i) APPLICANT: Choi et. al.  
 9 (ii) TITLE OF INVENTION: Streptococcus pneumoniae Antigens and  
 10 Vaccines  
 12 (iii) NUMBER OF SEQUENCES: 454  
 14 (iv) CORRESPONDENCE ADDRESS:  
 16 (A) ADDRESSEE: Human Genome Sciences, Inc.  
 17 (B) STREET: 9410 Key West Avenue  
 18 (C) CITY: Rockville  
 19 (D) STATE: Maryland  
 20 (E) COUNTRY: USA  
 21 (F) ZIP: 20850

## 24 (v) COMPUTER READABLE FORM:

- 26 (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 27 (B) COMPUTER: Dell Latitude C610  
 28 (C) OPERATING SYSTEM: Windows 2000  
 29 (D) SOFTWARE: ASCII Text

## 32 (vi) CURRENT APPLICATION DATA:

- C--> 34 (A) APPLICATION NUMBER: US/09/765,272A  
 C--> 35 (B) FILING DATE: 22-Jan-2001  
 36 (C) CLASSIFICATION:

## 39 (vii) PRIOR APPLICATION DATA:

- 41 (A) APPLICATION NUMBER: 08/961,083  
 42 (B) FILING DATE: OCT-30-1997

## 45 (viii) ATTORNEY/AGENT INFORMATION:

- 47 (A) NAME: Lin J. Hymel  
 48 (B) REGISTRATION NUMBER: 45,414  
 49 (C) REFERENCE/DOCKET NUMBER: PB340P2C2

## 52 (ix) TELECOMMUNICATION INFORMATION:

- 54 (A) TELEPHONE: (301) 610-5790  
 55 (B) TELEFAX: (301) 309-8439

## 58 (2) INFORMATION FOR SEQ ID NO: 1:

## 60 (i) SEQUENCE CHARACTERISTICS:

- 61 (A) LENGTH: 1999 base pairs  
 62 (B) TYPE: nucleic acid  
 63 (C) STRANDEDNESS: double  
 64 (D) TOPOLOGY: linear

## 67 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

69 TAAAATCTAC GACAATAAAA ATCAACTCAT TGCTGACTTG GGTTCCTGAAC GCCGCGTCAA	60
71 TGCCCAAGCT AATGATATTC CCACAGATTT GGTAAAGGCA ATCGTTTCTA TCGAAGACCA	120
73 TCGCTTCTTC GACCACAGGG GGATTGATAC CATCCGTATC CTGGGAGCTT TCTTGCAGCA	180
75 TCTGCAAAGC AATTCCCTCC AAGGTGGATC AACTCTCACC CAACAGTTGA TTAAGTTGAC	240

ENTERED

## RAW SEQUENCE LISTING

DATE: 02/20/2003

PATENT APPLICATION: US/09/765,272A

TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

```

77 TTACTTTTCA ACTTCGACTT CCGACCAGAC TATTTCTCGT AAGGCTCAGG AAGCTTGGTT      300
79 AGCGATTGAG TTAGAACAAA AAGCAACCAA GCAAGAAATC TTGACCTACT ATATAAATAA      360
81 GGTCTACATG TCTAATGGGA ACTATGGAAT GCAGACAGCA GCTCAAACT ACTATGGTAA      420
83 AGACCTCAAT AATTTAAGTT TACCTCAGTT AGCCTTGCTG GCTGGAATGC CTCAGGCACC      480
85 AAACCAATAT GACCCCTATT CACATCCAGA AGCAGCCCAA GACCGCCGAA ACTTGGTCTT      540
87 ATCTGAAATG AAAAATCAAG GCTACATCTC TGCTGAACAG TATGAGAAAG CAGTCAATAC      600
89 ACCAATTACT GATGGACTAC AAAGTCTCAA ATCAGCAAGT AATTACCCTG CTTACATGGA      660
91 TAATTACCTC AAGGAAGTCA TCAATCAAGT TGAAGAAGAA ACAGGCTATA ACCTACTCAC      720
93 AACTGGGATG GATGTCTACA CAAATGTAGA CCAAGAAGCT CAAAAACATC TGTGGGATAT      780
95 TTACAATACA GACGAATACG TTGCCTATCC AGACGATGAA TTGCAAGTCG CTTCTACCAT      840
97 TGTGATGTT TCTAACGGTA AAGTCATTGC CCAGCTAGGA GCACGCCATC AGTCAAGTAA      900
99 TGTTTCCTTC GGAATTAACC AAGCAGTAGA AACAAACCGC GACTGGGGAT CAACTATGAA      960
101 ACCGATCACA GACTATGCTC CTGCCTTGGA GTACGGTGTC TACGATTCAA CTGCTACTAT      1020
103 CGTTCACGAT GAGCCCTATA ACTACCCTGG GACAAATACT CCTGTTTATA ACTGGGATAG      1080
105 GGGCTACTTT GGCAACATCA CCTTGCAATA CGCCCTGCAA CAATCGCGAA ACGTCCCAGC      1140
107 CGTGGAACT CTAACAAGG TCGGACTCAA CCGCGCCAAG ACTTTCCTAA ATGGTCTAGG      1200
109 AATCGACTAC CCAAGTATTC ACTACTCAA TGCCATTTCA AGTAACACAA CCGAATCAGA      1260
111 CAAAAAATAT GGAGCAAGTA GTGAAAAGAT GGCTGCTGCT TACGCTGCCT TTGCAAATGG      1320
113 TGGAACCTAC TATAAACCAA TGTATATCCA TAAAGTCGTC TTTAGTGATG GGAGTGAAAA      1380
115 AGAGTTCTCT AATGTCGGAA CTCGTGCCAT GAAGGAAACG ACAGCCTATA TGATGACCGA      1440
117 CATGATGAAA ACAGTCTTGA CTTATGGAAC TGGACGAAAT GCCTATCTTG CTTGGCTCCC      1500
119 TCAGGCTGGT AAAACAGGAA CCTCTAATA TACAGACGAG GAAATTGAAA ACCACATCAA      1560
121 GACCTCTCAA TTTGTAGCAC CTGATGAAC ATTTGCTGGC TATACGCGTA AATATTCAAT      1620
123 GGCTGTATGG ACAGGCTATT CTAACCGTCT GACACCACTT GTAGGCAATG GCCTTACGGT      1680
125 CGCTGCCAAA GTTTACCGCT CTATGATGAC CTACCTGTCT GAAGGAAGCA ATCCAGAAGA      1740
127 TTGGAATATA CCAGAGGGGC TCTACAGAAA TGGAGAATTC GTATTTAAAA ATGGTGCTCG      1800
129 TTCTACGTGG AACTCACCTG CTCCACAACA ACCCCCATCA ACTGAAAGTT CAAGCTCATC      1860
131 ATCAGATAGT TCAACTTCAC AGTCTAGCTC AACCCTCCA AGCACAAATA ATAGTACGAC      1920
133 TACCAATCCT AACAATAATA CGCAACAATC AAATACAACC CCTGATCAAC AAAATCAGAA      1980
135 TCCTCAACCA GCACAACCA                                     1999

```

137 (2) INFORMATION FOR SEQ ID NO: 2:

139 (i) SEQUENCE CHARACTERISTICS:

140 (A) LENGTH: 666 amino acids

141 (B) TYPE: amino acid

142 (C) STRANDEDNESS: single

143 (D) TOPOLOGY: linear

145 (ii) MOLECULE TYPE: protein

148 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

150 Lys Ile Tyr Asp Asn Lys Asn Gln Leu Ile Ala Asp Leu Gly Ser Glu
151 1 5 10 15
153 Arg Arg Val Asn Ala Gln Ala Asn Asp Ile Pro Thr Asp Leu Val Lys
154 20 25 30
156 Ala Ile Val Ser Ile Glu Asp His Arg Phe Phe Asp His Arg Gly Ile
157 35 40 45
159 Asp Thr Ile Arg Ile Leu Gly Ala Phe Leu Arg Asn Leu Gln Ser Asn
160 50 55 60
162 Ser Leu Gln Gly Gly Ser Thr Leu Thr Gln Gln Leu Ile Lys Leu Thr
163 65 70 75 80
165 Tyr Phe Ser Thr Ser Thr Ser Asp Gln Thr Ile Ser Arg Lys Ala Gln

```

## RAW SEQUENCE LISTING

DATE: 02/20/2003

PATENT APPLICATION: US/09/765,272A

TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

166				85				90				95		
168	Glu	Ala	Trp	Leu	Ala	Ile	Gln	Leu	Glu	Gln	Lys	Ala	Thr	Lys
169				100				105				110		
171	Ile	Leu	Thr	Tyr	Tyr	Ile	Asn	Lys	Val	Tyr	Met	Ser	Asn	Gly
172				115				120				125		
174	Gly	Met	Gln	Thr	Ala	Ala	Gln	Asn	Tyr	Tyr	Gly	Lys	Asp	Leu
175				130				135				140		
177	Leu	Ser	Leu	Pro	Gln	Leu	Ala	Leu	Leu	Ala	Gly	Met	Pro	Gln
178				145				150				155		
180	Asn	Gln	Tyr	Asp	Pro	Tyr	Ser	His	Pro	Glu	Ala	Ala	Gln	Asp
181				165						170			175	
183	Asn	Leu	Val	Leu	Ser	Glu	Met	Lys	Asn	Gln	Gly	Tyr	Ile	Ser
184				180						185			190	
186	Gln	Tyr	Glu	Lys	Ala	Val	Asn	Thr	Pro	Ile	Thr	Asp	Gly	Leu
187				195				200				205		
189	Leu	Lys	Ser	Ala	Ser	Asn	Tyr	Pro	Ala	Tyr	Met	Asp	Asn	Tyr
190				210			215					220		
192	Glu	Val	Ile	Asn	Gln	Val	Glu	Glu	Glu	Thr	Gly	Tyr	Asn	Leu
193				225			230				235			240
195	Thr	Gly	Met	Asp	Val	Tyr	Thr	Asn	Val	Asp	Gln	Glu	Ala	Gln
196				245						250			255	
198	Leu	Trp	Asp	Ile	Tyr	Asn	Thr	Asp	Glu	Tyr	Val	Ala	Tyr	Pro
199				260				265					270	
201	Glu	Leu	Gln	Val	Ala	Ser	Thr	Ile	Val	Asp	Val	Ser	Asn	Gly
202				275				280					285	
204	Ile	Ala	Gln	Leu	Gly	Ala	Arg	His	Gln	Ser	Ser	Asn	Val	Ser
205				290			295					300		
207	Ile	Asn	Gln	Ala	Val	Glu	Thr	Asn	Arg	Asp	Trp	Gly	Ser	Thr
208				305			310				315			320
210	Pro	Ile	Thr	Asp	Tyr	Ala	Pro	Ala	Leu	Glu	Tyr	Gly	Val	Tyr
211				325						330			335	
213	Thr	Ala	Thr	Ile	Val	His	Asp	Glu	Pro	Tyr	Asn	Tyr	Pro	Gly
214				340					345				350	
216	Thr	Pro	Val	Tyr	Asn	Trp	Asp	Arg	Gly	Tyr	Phe	Gly	Asn	Ile
217				355			360					365		
219	Gln	Tyr	Ala	Leu	Gln	Gln	Ser	Arg	Asn	Val	Pro	Ala	Val	Glu
220				370			375					380		
222	Asn	Lys	Val	Gly	Leu	Asn	Arg	Ala	Lys	Thr	Phe	Leu	Asn	Gly
223				385			390				395			400
225	Ile	Asp	Tyr	Pro	Ser	Ile	His	Tyr	Ser	Asn	Ala	Ile	Ser	Ser
226				405					410				415	
228	Thr	Glu	Ser	Asp	Lys	Lys	Tyr	Gly	Ala	Ser	Ser	Glu	Lys	Met
229				420				425					430	
231	Ala	Tyr	Ala	Ala	Phe	Ala	Asn	Gly	Gly	Thr	Tyr	Tyr	Lys	Pro
232				435				440					445	
234	Ile	His	Lys	Val	Val	Phe	Ser	Asp	Gly	Ser	Glu	Lys	Glu	Phe
235				450			455					460		
237	Val	Gly	Thr	Arg	Ala	Met	Lys	Glu	Thr	Thr	Ala	Tyr	Met	Met
238				465			470				475			480

## RAW SEQUENCE LISTING

DATE: 02/20/2003

PATENT APPLICATION: US/09/765,272A

TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

```

240 Met Met Lys Thr Val Leu Thr Tyr Gly Thr Gly Arg Asn Ala Tyr Leu
241                               485                               490                               495
243 Ala Trp Leu Pro Gln Ala Gly Lys Thr Gly Thr Ser Asn Tyr Thr Asp
244                               500                               505                               510
246 Glu Glu Ile Glu Asn His Ile Lys Thr Ser Gln Phe Val Ala Pro Asp
247                               515                               520                               525
249 Glu Leu Phe Ala Gly Tyr Thr Arg Lys Tyr Ser Met Ala Val Trp Thr
250                               530                               535                               540
252 Gly Tyr Ser Asn Arg Leu Thr Pro Leu Val Gly Asn Gly Leu Thr Val
253                               545                               550                               555                               560
255 Ala Ala Lys Val Tyr Arg Ser Met Met Thr Tyr Leu Ser Glu Gly Ser
256                               565                               570                               575
258 Asn Pro Glu Asp Trp Asn Ile Pro Glu Gly Leu Tyr Arg Asn Gly Glu
259                               580                               585                               590
261 Phe Val Phe Lys Asn Gly Ala Arg Ser Thr Trp Asn Ser Pro Ala Pro
262                               595                               600                               605
264 Gln Gln Pro Pro Ser Thr Glu Ser Ser Ser Ser Ser Ser Asp Ser Ser
265                               610                               615                               620
267 Thr Ser Gln Ser Ser Ser Thr Thr Pro Ser Thr Asn Asn Ser Thr Thr
268                               625                               630                               635                               640
270 Thr Asn Pro Asn Asn Asn Thr Gln Gln Ser Asn Thr Thr Pro Asp Gln
271                               645                               650                               655
273 Gln Asn Gln Asn Pro Gln Pro Ala Gln Pro
274                               660                               665

```

276 (2) INFORMATION FOR SEQ ID NO: 3:

278 (i) SEQUENCE CHARACTERISTICS:

279 (A) LENGTH: 1714 base pairs

280 (B) TYPE: nucleic acid

281 (C) STRANDEDNESS: double

282 (D) TOPOLOGY: linear

286 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

288 AAATTACAAT ACGGACTATG AATTGACCTC TGGAGAAAAA TTACCTCTTC CTAAAGAGAT      60
290 TTCAGGTAC ACTTATATTG GATATATCAA AGAGGGAAAA ACGACTTCTG AGTCTGAAGT      120
292 AAGTAATCAA AAGAGTTCAG TTGCCACTCC TACAAAACAA CAAAAGGTGG ATTATAATGT      180
294 TACACCGAAT TTTGTAGACC ATCCATCAAC AGTACAAGCT ATTCAGGAAC AAACACCTGT      240
296 TTCTTCAACT AAGCCGACAG AAGTTCAAGT AGTTGAAAAA CCTTCTCTA CTGAATTAAT      300
298 CAATCCAAGA AAAGAAGAGA AACAATCTTC AGATTCTCAA GAACAATTAG CCGAACATAA      360
300 GAATCTAGAA ACGAAGAAAG AGGAGAAGAT TTCTCCAAAA GAAAAGACTG GGGTAAATAC      420
302 ATTAAATCCA CAGGATGAAG TTTTATCAGG TCAATTGAAC AAACCTGAAC TCTTATATCG      480
304 TGAGGAAACT ATGGAGACAA AAATAGATTT TCAAGAAGAA ATTCAAGAAA ATCTGATTT      540
306 AGCTGAAGGA ACTGTAAGAG TAAAACAAGA AGGTAAATTA GGTAAGAAAG TTGAAATCGT      600
308 CAGAAATATC TCTGTAAACA AGGAAGAAGT TTCGCGAGAA ATTGTTTCAA CTTCAACGAC      660
310 TGCGCCTAGT CCAAGAATAG TCGAAAAAGG TACTAAAAAA ACTCAAGTTA TAAAGGAACA      720
312 ACCTGAGACT GGTGTAGAAC ATAAGGACGT ACAGTCTGGA GCTATTGTTG AACCCGCAAT      780
314 TCAGCCTGAG TTGCCCGAAG CTGTAGTAAG TGACAAAGGC GAACCAGAAG TTCAACCTAC      840
316 ATTACCCGAA GCAGTTGTGA CCGACAAAGG TGAGACTGAG GTTCAACCAG AGTCGCCAGA      900
318 TACTGTGGTA AGTGATAAAG GTGAACCAGA GCAGGTAGCA CCGCTTCCAG AATATAAGGG      960
320 TAATATTGAG CAAGTAAAC CTGAAACTCC GGTGAGAAG ACCAAAGAAC AAGGTCCAGA     1020
322 AAAAAGTAA GAAGTTCCAG TAAAACCAAC AGAAGAAACA CCAGTAAATC CAAATGAAGG     1080

```

## RAW SEQUENCE LISTING

DATE: 02/20/2003

PATENT APPLICATION: US/09/765,272A

TIME: 16:02:52

Input Set : A:\340P2C2 Substitute Sequence List 14 Feb 2003.txt

Output Set: N:\CRF4\02202003\I765272A.raw

```

324 TACTACAGAA GGAACCTCAA TTCAAGAAGC AGAAAATCCA GTTCAACCTG CAGAAGAATC      1140
326 AACAAACGAAT TCAGAGAAAAG TATCACCAGA TACATCTAGC AAAAATACTG GGGAAAGTGC      1200
328 CAGTAATCCT AGTGATTCGA CAACCTCAGT TGGAGAATCA AATAAACCAG AACATAATGA      1260
330 CTCTAAAAAT GAAAAATCAG AAAAAACTGT AGAAGAAGTT CCAGTAAATC CAAATGAAGG      1320
332 CACAGTAGAA GGTACCTCAA ATCAAGAAAC AGAAAAACCA GTTCAACCTG CAGAAGAAAC      1380
334 ACAAACAAAC TCTGGGAAAA TAGCTAACGA AAATACTGGA GAAGTATCCA ATAAACCTAG      1440
336 TGATTCAAAA CCACCAAGTTG AAGAATCAAA TCAACCAGAA AAAAACGGAA CTGCAACAAA      1500
338 ACCAGAAAAT TCAGGTAATA CAACATCAGA GAATGGACAA ACAGAACCAG AACCATCAAA      1560
340 CGGAAATTCA ACTGAGGATG TTTCAACCGA ATCAAACACA TCCAATTCAA ATGGAAACGA      1620
342 AGAAATTTAA CAAGAAAATG AACTAGACCC TGATAAAAAG GTAGAAGAAC CAGAGAAAAAC      1680
344 ACTTGAATTA AGAAATGTTT CCGACCTAGA GTTA                                1714
346 (2) INFORMATION FOR SEQ ID NO: 4:
348     (i) SEQUENCE CHARACTERISTICS:
349         (A) LENGTH: 571 amino acids
350         (B) TYPE: amino acid
351         (C) STRANDEDNESS: single
352         (D) TOPOLOGY: linear
354     (ii) MOLECULE TYPE: protein
357     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
359     Asn Tyr Asn Thr Asp Tyr Glu Leu Thr Ser Gly Glu Lys Leu Pro Leu
360     1           5           10           15
362     Pro Lys Glu Ile Ser Gly Tyr Thr Tyr Ile Gly Tyr Ile Lys Glu Gly
363     20           25           30
365     Lys Thr Thr Ser Glu Ser Glu Val Ser Asn Gln Lys Ser Ser Val Ala
366     35           40           45
368     Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe
369     50           55           60
371     Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val
372     65           70           75           80
374     Ser Ser Thr Lys Pro Thr Glu Val Gln Val Val Glu Lys Pro Phe Ser
375     85           90           95
377     Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser
378     100          105          110
380     Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu
381     115          120          125
383     Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln
384     130          135          140
386     Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg
387     145          150          155          160
389     Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu
390     165          170          175
392     Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys
393     180          185          190
395     Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu
396     195          200          205
398     Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro
399     210          215          220
401     Arg Ile Val Glu Lys Gly Thr Lys Lys Thr Gln Val Ile Lys Glu Gln
402     225          230          235          240

```